Sử dụng phương pháp tương quan hạng trong nghiên cứu dân số

PHẠM QUỲNH HƯƠNG*

Khi nghiên cứu số liệu dân số, ta thấy giữa các dấu hiệu có mối liên hệ, ta nói rằng giữa chúng có mối tương quan. Vì là những dấu hiệu về lượng nên không những ta phát hiện được môi quan hệ phụ thuộc giữa chúng mà còn có thể lượng hóa được mối quan hệ ấy bằng hệ thức toán học cụ thể. Phương pháp tương quan được dùng để nghiên cứu môi liên hệ không hoàn toàn chặt chẽ được gọi là mối liên hệ phụ thuộc thống kê. Nghĩa là một tập hợp các giá trị của dấu hiệu này tương ứng với một giá trị của dấu hiệu khác dưới dạng phân bố thống kê.

Thông thường một dấu hiệu nào đó chịu ảnh hưởng của một vài dấu hiệu khác, trong đó có vài dấu hiệu đáng kể. Ta sẽ xét đến những dấu hiệu có ý nghĩa nhất định để xác định mối liên hệ. Còn những dấu hiệu khác được coi như không thay đổi. Những dấu hiệu chọn ra thường có dấu hiệu nguyên nhân và dấu hiệu kết quả hay dấu hiệu phụ thuộc. Với những dấu hiệu có thể xếp hạng theo thang thứ tự, ta có thể tính hệ số tương quan hạng. Trong bài nhỏ viết này tôi muốn trình bày hai phương pháp tính hệ số tương quan[1].

Hệ số tương quan hạng Spearman

Giả sử ta có hai biến x và y với x là biến độc lập và y là biến phụ thuộc. Hệ số Spearman được tính như sau: sắp xếp các phần từ thành các hạng theo thứ tự (tảng hoặc giảm dẫn đối với biến x, ta sẽ được các hạng Rx. Tiếp theo, sắp xếp tương ứng biến y ta sẽ được các hạng Ry. Sự chênh lệch giữa hai hạng Rx và Ry là d; =[Rx- Ry]. Hệ số Searman được tính

$$S = 1 - \frac{6\sum d_i^2}{n(n+1)(n-1)}$$

Ở đây, n là số phần tử.

Giá trị của2s thay đổi từ - 1 đến + 1 . S = + 1 khi tất cả các hạng Rx và Ry là như nhau, khiến cho $d^2 = 0$. Khi đó quan hệ x và y là thuận hoàn toàn. Ngược lại S= -1 là quan hệ nghịch hoàn toàn và S= 0 khi không có quan hệ giữa hai biến.

Ta hãy xét ví dụ:

Phụ nữ ở những độ tuổi khác nhau có hiểu biết về vòng tránh thai khác nhau. Ảnh .

---

* Cán bô phòng phương pháp và kỹ thuật Xã hội học, Viện Xã hội học.

[1] Các số liệu trong bài này là kết qua điều tra của Trung ương Hội liên hiệp phụ nữ Việt Nam tại ba tỉnh Quảng Nam, Hà Nội và Hải Hưng năm 1987

hưởng của độ tuổi tới việc hiểu biết về vòng tránh thai như thế nào? Ta cố thể sử dụng hệ số trung quan hạng Spearman để đánh giá mức độ tương quan.

Bảng 1: Số người có hiểu biết về vòng tránh thai phân theo độ tuổi.

Tuổi (x)	Hạng (Rx)	Hiểu biết (y)	Hạng (Ry)	di	di^2
20- 24	1	93,33	2	1	1
5- 29	2	99,22	5	3	9
30- 34	3	98,62	4	1	1
5- 39	4	57,81	1	3	9
40- 44	5	99,37	6	1	1
45+	6	97,43	3	3	9
					30

Trong bảng trên biến tuổi là biến độc lập, biến x; biến sự hiểu biết là biến y, biến phụ thuộc. Sự biến thiên của biến này ứng với sự biến thiên của biến thiên của biến kia. Hệ số Spaerman được tính dựa trên sự sắp hạng của số liệu. Cột hạng (Rx) là sự sắp xếp độ tuổi theo thứ 1 là tăng dần. Cột di được tính theo công thức nêu trên. Theo công thức đó ta có

$$S = 1 - \frac{6\sum d_i^2}{n(n+1)(n-1)} = 1 - \frac{6.30}{6.7.5} = 0,13$$

Hệ số S > 0, như vậy sự tương quan giữa độ tuổi và sự hiểu biết là tương quan thuận, tức là độ tuổi càng cao sự hiểu biết càng nhiều. Nhưng sự tương quan này là thấp, vì giá trị của hệ số S nhỏ.

Hệ số Kendall:

Khi ta có hai dấu hiệu x và y mà giá trị của dấu hiệu này tương ứng với một tập hợp giá trị của dấu hiệu kia dưới dạng phân bố thống kê, ta có thể sử dụng hệ số Kendall để đánh giá tương quan. Giả sử x là biến độc lập và biến y là biến phụ thuộc. Với mỗi giá trị của x ta có một tập hợp các giá trị của y. Trước hết, với mỗi giá trị của x ta sấp hạng tăng hoặc giâm dần.

Gọi Rj là tổng các hạng cho từng phần tử của y.

Rj là số trung bình của tổng các hạng này.

Hệ số Kendall được tính như sau:

hệ số Spea "man:

$$W = \frac{S^2}{1/12\ K^2 (n^2-n)}$$

Ở đây, n là số phần tử của x

k là số phần tử của y

$$R = \sum_{1}^{n} Rj / n \qquad S_j^2 = \sum_{j=1}^{K} (Rj - R)$$

W lấy giá trị trong khoảng 0,1.

Xét ví dụ sau:

Bảng 2: Số con phân theo trình độ văn hóa của mẹ

Văn hóa	Số con			
	0	1	2	3+
Cấp I	1,8	0,01	38,6	47,4
Cấp II	0,3	24,9	43,7	3l,1
Cấp III	1,6	35,7	39,3	23,3
Đại hoc	4,5	33,1	45,4	17,0

Ở đây ta coi biến văn hóa là biến độc lập và số con là biến phụ thuộc. Trong từng hàng (từng cấp văn hóa) ta sắp xếp tỷ lệ số con theo thứ tự tăng dần, ta được bảng sau:

Bảng xếp hạng

Văn hóa	Số con			
Cấp I	0	1	2	3+
Cấp II	2	1	3	4
Cấp III	1	2	4	3
Đại học	1	3	4	2
Rj	1	3	4	2
$Rj^2=(Rj-R)^2$	5	9	15	11
	25	1	25	1

Rj: tổng theo từng cột của các hạng được xếp.

Rj tổng các Rj =40 (J =l)

Trung bình cộng các Rj là: R = Rj / số cột = 40 /4 = 10 Ta có hệ tương quan Kendall:

$$W = \frac{Sj^2}{1/12\ K^2 (n^2-n)} = \frac{52}{1/12.\ 4^2 (4^3-4)} = 0,65$$

với n là số hàng và k là số cột.

Với W = 0, 65 mối tương quan giữa trình độ văn hóa và số con là tương đối chặt.

Từ những hệ số tương quan Spearman, Kendall ta có thể rút ra được những nhận xét về mối tương quan giữa các biến. Nhưng đây chỉ là những nhận xét, những đánh giá rút ra từ những hệ số tương quan. Trên cơ sở những nhận xét, đánh giá này ta có thể đưa ra những giả thuyết thống kê. Muốn xác định được sự đúng đắn của những giả thuyết đó người ta phải dùng tới những phương pháp kiểm định giả thuyết. Vê mặt lý thuyết những hệ số tương quan được tính với n nhỏ hơn 20 đều bắt buộc phải có bước kiểm định già thuyết.